

**Gaussianity of LISA's confusion backgrounds**Étienne Racine<sup>1</sup> and Curt Cutler<sup>2</sup><sup>1</sup>*Department of Physics, Mathematics and Astronomy, California Institute of Technology, Pasadena, California 91125, USA*<sup>2</sup>*Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109, USA*

(Received 30 August 2007; published 27 December 2007)

Data analysis for the proposed Laser Interferometer Space Antenna (LISA) will be complicated by the huge number of sources in the LISA band. In the frequency band  $\sim 10^{-4} - 2 \times 10^{-3}$  Hz, galactic white dwarf binaries (GWDBs) are sufficiently dense in frequency space that it will be impossible to resolve most of them, and “confusion noise” from the unresolved Galactic binaries will dominate over instrumental noise in determining LISA's sensitivity to other sources in that band. Confusion noise from unresolved extreme-mass-ratio inspirals (EMRIs) could also contribute significantly to LISA's total noise curve. To date, estimates of the effect of LISA's confusion noise on matched-filter searches and their detection thresholds have generally approximated the noise as Gaussian, based on the central limit theorem. However in matched-filter searches, the appropriate detection threshold for a given class of signals may be located rather far out on the tail of the signal-to-noise probability distribution, where *a priori* it is unclear whether the Gaussian approximation is reliable. Using the Edgeworth expansion and the theory of large deviations, we investigate the probability distribution of the usual matched-filter detection statistic, far out on the tail of the distribution. We apply these tools to four somewhat idealized versions of LISA data searches: searches for EMRI signals buried in GWDB confusion noise, and searches for massive black hole binary signals buried in (i) GWDB noise, (ii) EMRI noise, and (iii) a sum of EMRI noise and Gaussian noise. Assuming reasonable short-distance cutoffs in the populations of confusion sources (since the very closest and hence strongest sources will be individually resolvable), modifications to the appropriate detection threshold, due to the non-Gaussianity of the confusion noise, turn out to be quite small for realistic cases. The smallness of the correction is partly due to the fact that these three types of sources evolve on quite different time scales, so no single background source closely resembles any search template. We also briefly discuss other types of LISA searches where the non-Gaussianity of LISA's confusion backgrounds could perhaps have a much greater impact on search reliability and efficacy.

DOI: [10.1103/PhysRevD.76.124033](https://doi.org/10.1103/PhysRevD.76.124033)

PACS numbers: 04.25.Nx, 04.30.Db, 04.80.Nn, 95.75.Wx

**I. INTRODUCTION**

Data analysis for the proposed Laser Interferometer Space Antenna (LISA) will be complicated by the huge number of sources in the LISA band. For example, while LISA is expected to detect of the order of  $10^4$  individual compact binaries (especially white dwarf–white dwarf binaries) in our Galaxy, in the frequency band  $\sim 10^{-4} - 2 \times 10^{-3}$  Hz such binaries are sufficiently dense in frequency space that it will be impossible to resolve most of them. The “confusion noise” from all the unresolved Galactic binaries will dominate over instrumental noise in determining LISA's sensitivity to other sources in that band. Extreme-mass-ratio inspirals (EMRIs) are another very important category of LISA sources. EMRIs are inspirals of stellar-mass compact objects (white dwarfs, neutron stars, or black holes) into massive ( $\sim 10^6 M_\odot$ ) black holes (MBHs) in galactic nuclei. Because of their extremely small mass ratio, EMRI sources remain in the LISA band for time scales of the order of years. While LISA can do a great deal of interesting science with individually detected EMRIs, the EMRIs that are too faint to be resolved also constitute a confusion background, partially masking other sources. Barack and Cutler [1] (hereinafter BC) estimated the spectral density of confusion noise from unresolved

EMRIs and found that it becomes comparable to that of LISA's instrumental noise or white dwarf (WD) confusion noise only if EMRI event rates turn out to be at the high end of the estimated range. BC therefore concluded that LISA's EMRI confusion background would be rather benign: either the EMRI rates are low-to-medium, in which case non-EMRI noise sources dominate the total noise, or the EMRI rates are high, in which case noise from unresolvable EMRIs could dominate, but the EMRI detection rate is *also* higher (which would more than compensate, from a scientific standpoint).

However there is a potential caveat to BC's treatment of EMRI confusion noise (as well as to many discussions of the white dwarf confusion noise) related to the Gaussianity of that noise. Because the number of undetected galactic white dwarf binaries (GWDBs) or EMRIs will be large ( $\sim 10^7$ – $10^8$  for GWDBs and  $\sim 10^5$ – $10^6$  for EMRIs), the confusion background has generally been treated as approximately Gaussian, based on an appeal to the central limit theorem. However in matched-filter searches, the detection threshold for a given class of signals may be located rather far out on the tail of the signal-to-noise probability distribution. For example, in searching for EMRIs, the vast number of independent EMRI signals that can be searched-for necessitates a detection threshold of  $\sim 14\sigma$

[2], assuming Gaussian statistics; similarly we estimate that searches for massive black hole binaries (MBHBs) will require a signal-to-noise detection threshold of  $\sim 7\sigma$ , to keep false alarms at an acceptable level. Then naturally one must confront the question: How much does the tail of the distribution of the usual detection statistic deviate from Gaussian in the range  $\sim 7 - 14\sigma$ ? In other words, how often does the confusion background manage to mimic the signals searched for, at the level of the usual detection threshold? Or put yet another way: how much *higher* must one set the detection threshold to compensate for the non-Gaussianity of the tail of the distribution? We analyze these questions by harnessing two tools from statistics, namely, the Edgeworth expansion and the theory of large deviations, and applying them to model problems that are somewhat idealized versions of the cases that will arise in actual LISA data analysis.

In this paper we will be concerned with three types of LISA sources, all of which are binaries: GWDBs, MBHBs, and EMRIs. To make the calculations below analytically tractable, we shall assume that the binary orbits are quasi-circular (i.e., circular except for a slow inspiral due to gravitational radiation reaction), and we shall approximate each gravitational waveform by its lowest-order piece in a post-Newtonian expansion. Also, while LISA should return two independent science data channels (and a third at high frequency), for simplicity we shall treat the output as a single channel. Other simplifications and approximations are discussed below.

To avoid confusion, we should emphasize that in this paper we are mainly interested in detections near the threshold signal-to-noise ratio (SNR). Now, the strongest MBHB signals detected by LISA (perhaps from  $\sim 10^6 M_\odot$  MBHBs merging at redshift  $z < 1$ ) will likely have (matched-filtering) SNRs  $\sim 10^3$ ; however those are *not* the MBHB signals that interest us here. Instead, when we discuss searches for MBHB signals, we are mainly interested in the most distant resolvable ones; e.g., mergers of  $\sim 10^4 M_\odot$  MBHBs at  $z \sim 20$ . Likewise when we discuss searches for EMRI signals, our interest is in the weakest resolvable ones. Of course, the overall detection rates will likely be dominated by the weakest detectable signals.

This paper is structured as follows. We first discuss, in Sec. II, the role of confusion noise in a matched-filter search and show how it reduces to a statistical problem involving sums of independent identically distributed random variables. We then describe in some detail the two statistical techniques we apply in this paper, namely, the Edgeworth expansion and the theory of large deviations. Next, in Sec. III, we describe in detail our toy models for the GWDB and EMRI confusion noise, respectively. In particular we discuss our model waveforms, which are simply Newtonian circular-orbit chirps, and our choices for binary parameter distributions. Finally in Sec. IV we use the confusion noise models of Sec. III and apply the

tools described in Sec. II to four model searches: searches for EMRI signals imbedded in GWDB confusion noise and searches for MBHBs imbedded in (i) GWDB noise, (ii) EMRI noise, and (iii) the sum of EMRI noise and Gaussian noise (instrumental plus GWDB), respectively. In each case we obtain the probability distributions for the usual detection statistic and assess the impact of the non-Gaussianity of the confusion noise on the appropriate detection threshold. Our conclusions are summarized in Sec. V. In an appendix we present a heuristic derivation of the central result in large-deviations theory (Chernoff's formula), describe its relation to the Edgeworth expansion, and apply it to a simple, illustrative case—the binomial distribution.

Throughout this paper, we use geometrical units in which  $G = c = 1$ . Therefore everything can be measured in our fundamental unit of seconds. For familiarity, we sometimes express quantities in terms of year, Mpc, or  $M_\odot$ , which are related to our fundamental unit by  $1 \text{ yr} = 3.1556 \times 10^7 \text{ s}$ ,  $1 \text{ Mpc} = 1.029 \times 10^{14} \text{ s}$ , and  $1 M_\odot = 4.926 \times 10^{-6} \text{ s}$ .

## II. STATISTICAL FOUNDATIONS

### A. Confusion noise in matched-filter searches

As an introduction to the general problem of searching for gravitational-wave (GW) signals that may be buried in confusion noise, consider a LISA data set  $s(t)$  that is dominated by instrumental noise plus unresolved background signals, but which may also contain some resolvable signal proportional to  $h(t)$ ; i.e.,

$$s(t) = n(t) + \sum_{i=1}^N h_i(t) + \rho h(t) \quad (2.1)$$

where  $n(t)$  represents (Gaussian) instrumental noise, the sum over  $N$  sources represents the confusion background,  $h(t)$  is the sought-for signal [normalized to  $\langle h|h \rangle = 1$ , where the inner product  $\langle \cdot | \cdot \rangle$  is defined below], and  $\rho$  represents the overall strength of the sought-for signal. If  $h(t)$  is simply not present in the data, then  $\rho = 0$ . For example, the background signals  $h_i(t)$  could be from GWDBs,<sup>1</sup> while  $h(t)$  is the gravitational-wave signal from some MBHB. In a matched-filter search for  $h(t)$  in this data set, one basically just computes the inner product  $\langle s|h \rangle$ :

$$\langle s|h \rangle = \langle n|h \rangle + \sum_{i=1}^N \langle h_i|h \rangle + \rho, \quad (2.2)$$

where  $\langle n|h \rangle$  is a Gaussian random variable, and where each  $x_i \equiv \langle h_i|h \rangle$  is a random variable drawn from some probability distribution function (PDF)  $p(x)$ . There must be

<sup>1</sup>Since the WD binaries are located much closer to us than the sought-for MBHB, their summed signal is sometimes called “foreground” instead of “background.” However we shall refer to all confusion noise populations simply as “backgrounds.”

some threshold value  $\rho_{\text{th}}$ , such that when  $(s|h) > \rho_{\text{th}}$  one can claim a detection with very high confidence (say, >99%). But what is this threshold value? To compute  $\rho_{\text{th}}$ , we need to know the probability distribution of the sum

$$X \equiv \sum_i^N x_i. \quad (2.3)$$

Most of the work in this paper will be spent in estimating the probability distribution function for  $X$ ,  $P_N(X)$ , given its parent distribution  $p(x)$ . We will be particularly concerned with the behavior of  $P_N(X)$  at large  $X$ —out on the “high- $\sigma$ ” tail. In the next subsection we describe two tools from statistics that are quite useful in this context.

### B. The central limit theorem and beyond

Let  $p(x)$  be some normalized PDF. The  $q$ th moment of  $p(x)$  is defined to be

$$\mu_q \equiv E[x^q] = \int x^q p(x) dx. \quad (2.4)$$

We shall assume for convenience that  $\mu_1$  (the mean value of  $x$ ) vanishes, since it automatically does so in all applications in this paper. We shall also assume that the second moment  $\mu_2$  (the variance of  $x$ ) exists, and define  $\sigma_x \equiv \sqrt{\mu_2}$ . Let  $Z$  be the average value of  $N$  samples from this distribution:

$$Z \equiv \frac{1}{N} \sum_{i=1}^N x_i. \quad (2.5)$$

Then the central limit theorem basically states that in the limit of large  $N$ , the PDF for  $Z$ ,  $P_N(Z)$ , approaches a Gaussian with variance  $\mu_2/N$ . Defining the rescaled variable  $Y \equiv \frac{\sqrt{N}}{\sigma_x} Z$ , we have

$$P_N(Y) \rightarrow \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{Y^2}{2}\right] \quad (2.6)$$

as  $N \rightarrow \infty$ .

While the central limit theorem states that  $P_N(Y)$  converges to a Gaussian for large  $N$ , for this paper it is crucial to realize that the convergence of the ratio  $P_N(Y)/[(2\pi)^{-1/2}e^{-Y^2/2}]$  to unity can be remarkably slow at large values of  $Y$ . This is particularly true if some higher moments of  $p(x)$  diverge, as happens, e.g., if  $p(x)$  has only power-law decrease at large  $x$ .

To quote standard theorems on the convergence of  $P_N(Y)$  to a Gaussian, we need a few more definitions. Define  $F_N(Y)$  to be the cumulative distribution function (CDF) of  $P_N(Y)$ ,

$$F_N(Y) = \int_{-\infty}^Y P_N(\tilde{Y}) d\tilde{Y}, \quad (2.7)$$

and let  $\bar{F}_N(Y)$  be the complementary function to  $F_N(Y)$ :

$$\bar{F}_N(Y) = \int_Y^{\infty} P_N(\tilde{Y}) d\tilde{Y} = 1 - F_N(Y). \quad (2.8)$$

Also define  $\Phi(Y)$  to be the CDF of a Gaussian,

$$\Phi(Y) \equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^Y e^{-\tilde{Y}^2/2} d\tilde{Y}. \quad (2.9)$$

and let

$$\bar{\Phi}(Y) \equiv \frac{1}{\sqrt{2\pi}} \int_Y^{\infty} e^{-\tilde{Y}^2/2} d\tilde{Y} = 1 - \Phi(Y). \quad (2.10)$$

Of course,  $\bar{\Phi}(Y) = (1/2)\text{erfc}(Y/\sqrt{2})$ , where “erfc” is the complementary error function.

Now let us further assume that the absolute third moment  $\rho_3 \equiv E[|x|^3] > 0$  of the parent distribution  $p(x)$  exists and is finite. For this case, a well-known result on the convergence of (2.6) is the Berry-Esséen theorem, which states that for all  $Y$  and  $N$ ,

$$\sup_Y |F_N(Y) - \Phi(Y)| \leq C \frac{\rho_3}{\sigma_x^3} N^{-1/2}. \quad (2.11)$$

where  $C$  is some constant less than 0.7655 [3,4]. Of course, this is equivalent to

$$\sup_Y |\bar{F}_N(Y) - \bar{\Phi}(Y)| \leq C \frac{\rho_3}{\sigma_x^3} N^{-1/2}. \quad (2.12)$$

Now let us consider the practical implications of the Berry-Esséen theorem. What threshold value  $Y_{\text{th}}$  ensures that, say,  $\bar{F}_N(Y_{\text{th}}) < 10^{-6}$ ? Since  $\bar{\Phi}(4.8916) = 10^{-6}$ , a first estimate based on the central limit theorem would be  $Y_{\text{th}} \approx 4.8916$ . However, by Eq. (2.12), the error in this estimate (for  $\rho_3/\sigma_x^3$  of order one) can be of order  $N^{-1/2}$ . So the potential error in the Gaussian estimate greatly exceeds that estimate itself unless  $N > 10^{12}$ ! More generally, for large  $Y$ ,  $N$  must be exponentially large—of order  $e^{Y^2}$ —for the right-hand side of (2.12) to be smaller than  $\bar{\Phi}(Y)$ .

When higher moments of  $p(x)$  exist, one can systematically improve on the central limit theorem estimate of  $P_N(Y)$ . These improvements are described in the next two subsections.

### C. The Edgeworth expansion

The key ingredient in constructing the Edgeworth expansion is the cumulant generating functional of a PDF, defined as

$$\lambda(\omega) \equiv \ln E[e^{i\omega x}] = \ln \int_{-\infty}^{\infty} e^{i\omega x} p(x) dx. \quad (2.13)$$

One can expand the exponential and then the logarithm about  $\omega = 0$  in (2.13) to obtain the following series for the cumulant generating functional:

$$\lambda(\omega) = \sum_{q=2}^{\infty} \frac{\kappa_q}{q!} (i\omega)^q, \quad (2.14)$$

where  $\kappa_q$  is called the  $q$ th cumulant of the parent distribution. Now consider the cumulant generating functional  $\Lambda(\omega)$  of  $P_N(Y)$ , which is given by

$$\begin{aligned}\Lambda(\omega) &= \ln E[e^{i\omega Y}] = N\lambda\left(\frac{\omega}{\sqrt{N\sigma_x^2}}\right) \\ &= N \sum_{q=2}^{\infty} \frac{\kappa_q/\sigma_x^q}{q!} \left(\frac{i\omega}{\sqrt{N}}\right)^q,\end{aligned}\quad (2.15)$$

where  $\sigma_x^2 = \kappa_2$ . Notice that  $\Lambda(\omega)$  depends only on  $N$  and the cumulants of  $p(x)$ . Taking the exponential of both sides of (2.15), formally expanding the results around  $\omega = 0$ , and then gathering terms according to powers of  $N^{-1/2}$  yields

$$E[e^{i\omega Y}] = e^{-\omega^2/2} \left[ 1 + \sum_{r=1}^{\infty} \frac{P_r(i\omega)}{N^{r/2}} \right], \quad (2.16)$$

where  $P_r(i\omega)$  is a polynomial in  $i\omega$  depending only on the cumulants  $\kappa_q$ . Since the left-hand side of (2.16) is simply the Fourier transform of  $P_N(Y)$ , taking the inverse transform on both sides of (2.16) finally gives

$$P_N(Y) = \frac{1}{\sqrt{2\pi}} e^{-Y^2/2} \left[ 1 + \sum_{r=1}^{\infty} \frac{Q_r(Y)}{N^{r/2}} \right], \quad (2.17)$$

where the  $Q_r(Y)$  are polynomials in  $Y$ . The first few terms of the series are

$$\begin{aligned}P_N(Y) &= \frac{1}{\sqrt{2\pi}} e^{-Y^2/2} \left[ 1 + \frac{\kappa_3/\sigma_x^3}{6\sqrt{N}} H_3(Y) + \frac{\kappa_4/\sigma_x^4}{24N} H_4(Y) \right. \\ &\quad \left. + \frac{\kappa_3^2/\sigma_x^6}{72N} H_6(Y) + O(N^{-3/2}) \right],\end{aligned}\quad (2.18)$$

where  $H_q(Y)$  is the Chebyshev-Hermite polynomial of order  $q$ , the ones appearing above being

$$H_3(Y) = Y^3 - 3Y \quad (2.19a)$$

$$H_4(Y) = Y^4 - 6Y^2 + 3 \quad (2.19b)$$

$$H_6(Y) = Y^6 - 15Y^4 + 45Y^2 - 15. \quad (2.19c)$$

At this point we should emphasize that while the Edgeworth series (2.17) is formally correct, it does not converge in general. Rather, in the limit  $N \rightarrow \infty$  it provides an asymptotic expansion of the true CDF  $F_N(Y)$ . More precisely, let  $\Lambda_r(Y)$  be related to the polynomials  $Q_r(Y)$  defined above by

$$\Lambda_r(Y) = \int_{-\infty}^Y \frac{e^{-\tilde{Y}^2/2}}{\sqrt{2\pi}} Q_r(\tilde{Y}) d\tilde{Y}. \quad (2.20)$$

Assume that the first  $k$  cumulants  $\kappa_q$  exist, for some  $k \geq 3$ . Also assume that  $\lim_{T \rightarrow \infty} \sup_{|t| > T} |\nu(t)| < 1$ , where

$$\nu(t) \equiv \int e^{itx} p(x) dx. \quad (2.21)$$

[This condition on  $\nu(t)$  will be easily satisfied for all parent distributions  $p(x)$  we consider.] Then Theorem 3 in Sec. IV of Petrov [5] states that

$$\begin{aligned}\lim_{N \rightarrow \infty} N^{(k-2)/2} (1 + |Y|^k) \\ \times \left| F_N(Y) - \Phi(Y) - \sum_{r=1}^{k-2} N^{-r/2} \Lambda_r(Y) \right| = 0\end{aligned}\quad (2.22)$$

uniformly in  $Y$  ( $-\infty < Y < \infty$ ).

Assuming that  $p(-x) = p(x)$  (as will be true for all examples considered in this paper), so that the odd cumulants of  $p(x)$  all vanish, and assuming the first  $k$  cumulants exist (for  $k$  even), then this theorem implies that the error in the  $(k-2)$ th-order approximation to  $F_N(Y)$  scales like  $N^{-k/2}$  as  $N \rightarrow \infty$ . For example, assuming  $\kappa_4$  exists, the error in the second-order Edgeworth expansion of  $F_N(Y)$  scales like  $N^{-2}$  for sufficiently large  $N$ . E.g., assuming  $\kappa_4/\sigma_x^4$  is of order 1, one therefore generally requires  $N > 10^8$  for this potential error to be smaller than  $10^{-16}$ . (Again, if one uses  $\sim 10^{14}$  independent templates in the search, then one would want the false alarm probability for any one of them to be smaller than  $\sim 10^{-16}$ .)

Now fix  $Y$  and  $N$ . Since the Edgeworth series is only asymptotic, one will typically find that the first few terms in the series might get smaller and smaller, and their sum ever closer to  $P_N(Y)$ , but eventually the terms in the series may start to grow and the sum diverges. A useful rule of thumb is then to truncate the Edgeworth expansion before the first term that is larger than the previous ones.

If not all moments  $\kappa_q$  exist, it becomes clear why the Edgeworth series cannot converge, since all terms in the expansion decrease exponentially with  $Y$  at large  $Y$ , while  $F_N(Y)$  falls off much more slowly. To see this, consider the case where  $p(x)$  is an even function having a power-law tail:

$$p(x) \rightarrow B\sigma_x^m x^{-m-1} \quad \text{for } |x| \gg \sigma_x \quad (2.23)$$

for some constant  $B$  and some odd  $m > 0$ . Let  $f(x)$  be the CDF for  $p(x)$ , and let  $\tilde{f}(x) \equiv 1 - f(x)$ . Then clearly  $\tilde{f}(x) \rightarrow (B/m)(x/\sigma_x)^{-m}$  at large  $x$ . Now fix the number of samples,  $N$ . Following Bazant [6], we note that the probability that the sum  $\sum_{i=1}^N (N^{1/2}\sigma_x)^{-1} x_i$  is greater than some value  $Y$  is clearly of the same order or greater than the probability that any single term in the sum is greater than  $Y$ , so

$$\bar{F}_N(Y) \geq N \tilde{f}(N^{1/2}\sigma_x Y) \rightarrow (B/m) N^{1-m/2} Y^{-m} \quad (2.24)$$

at large  $Y$ . So if the parent distribution has a power-law tail, then for any fixed  $N$ ,  $P_N(Y)$  has the same power-law falloff at very large  $Y$ . (Of course, the above argument just shows that  $\bar{F}_N(Y)$  falls off no faster than  $Y^{-m}$ ; however it seems likely that  $\bar{F}_N(Y)$  and  $\tilde{f}(x)$  fall off according to the same power law at large  $Y$  and  $x$ , respectively [6].)



This line of reasoning suggests that the Edgeworth expansion becomes unreliable at values of  $Y$  such that

$$(B/m)N^{1-m/2}Y^{-m} > Y^{-1}e^{-Y^2/2} \quad (2.25)$$

or

$$Y^2 > (m-2) \ln N + 2(m-1) \ln Y - 2 \ln(B/m). \quad (2.26)$$

We shall typically be interested in cases where  $N$  is large enough that the  $(m-2) \ln N$  dominates the right-hand side of (2.26). In that case, we obtain the rule of thumb that the Edgeworth expansion (and its first term, the central limit theorem estimate) become unreliable for  $Y > \sqrt{(m-2) \ln N}$ . So the range of validity of the Edgeworth expansion increases only like the square root of the exponent  $m$  describing the power-law falloff of  $p(x)$ .

The situation changes dramatically if the parent distribution  $p(x)$  falls to zero exponentially (or faster) as  $x \rightarrow \infty$ . In that case large-deviations theory guarantees that  $P_N(Z)$  also has exponential falloff as  $Z \rightarrow \infty$ . We turn to this subject next.

#### D. The theory of large deviations

The goal of large-deviations theory is to determine the PDF of the random variable  $Z$ , defined above in Eq. (2.5), on the high- $\sigma$  tails. From the parent distribution  $p(x)$ , one begins by defining a modified cumulant generating functional  $\lambda(\beta)$  as

$$\lambda(\beta) = \ln \int e^{\beta x} p(x) dx. \quad (2.27)$$

[Comparing with (2.13), we see that this is simply the usual cumulant generating functional evaluated at imaginary frequency  $\omega = -i\beta$ .] Note that this integral does not exist unless  $p(x)$  falls to zero exponentially fast as  $x \rightarrow \infty$ . To emphasize this point: if  $p(x)$  has a power-law tail as  $x \rightarrow \infty$ , then  $\lambda(\beta)$  does not exist and the results of large-deviations theory do not apply. In the rest of this subsection, we will assume  $p(x)$  is sufficiently well behaved at large  $x$  that  $\lambda(\beta)$  exists. Then the basic result of large-deviations theory is a theorem due to Cramér (e.g., see [7]), which states that

$$\lim_{N \rightarrow \infty} \frac{1}{N} \ln \bar{F}_N(Z) \rightarrow -I(z), \quad (2.28)$$

where

$$I(Z) = \max_{\beta} [Z\beta - \lambda(\beta)]. \quad (2.29)$$

This basically implies that for large  $N$  and arbitrary  $Z$ , the PDF of the random variable  $Z$  is well approximated by

$$P_N(Z) \approx C \exp[-NI(Z)] \quad (2.30)$$

where  $C$  is a normalization constant. Equation (2.29) is sometimes referred to in the literature as Chernoff's for-

mula, and the function  $I(Z)$  is called the "rate function." Clearly  $I(Z)$  is the Legendre transform of  $\lambda(\beta)$ .

There are well-known, close connections between Chernoff's formula and statistical mechanics. Roughly,  $\beta$  is like an inverse-temperature,  $\lambda(\beta)$  is analogous to the Helmholtz free energy, and  $-NI(Z)$  is analogous to the entropy. A gentle introduction to large-deviations theory is given in Ref. [8]. Since we presume most of our readers are unfamiliar with large-deviations theory, we also give a short, heuristic derivation of Eqs. (2.29) and (2.30) in Appendix A.

### III. CONFUSION NOISE FROM POPULATIONS OF BINARIES

In this paper, all the GW sources we consider are types of binaries: WD binaries, MBHBs, and EMRIs. We shall be considering the problems of searching for one type of binary in the confusion noise produced by a large number of unresolved sources of a different type; e.g., considering the search for MBHBs buried in the confusion background of unresolved WDs or unresolved EMRIs. Since this paper represents a first-cut at the problem of estimating the non-Gaussian tails of the detection statistic, we shall simplify the analysis by approximating all three types of binaries as being in (nonprecessing) quasicircular orbits. We further approximate the emitted gravitational waveform as a simple chirp, with instantaneous frequency  $f$  equal to twice the orbital frequency. Also, while the waveform that LISA actually measures is modulated (on a 1-year time scale) by the satellite constellation's rotational and translational motion, for simplicity we neglect these modulations. It should be clear from the derivations, however, that including LISA's orbital modulations would have very little impact on our basic results. In the next subsection we briefly describe our model gravitational waveforms and their overlaps.

#### A. The waveforms from circular-orbit binaries and their overlaps

Using the quadrupole formula, one can show that (to lowest order in a post-Newtonian expansion) the instantaneous gravitational-wave frequency  $f$  evolves in time according to

$$\dot{f} = \frac{96}{5} \pi^{8/3} \mu M^{2/3} f^{11/3}. \quad (3.1)$$

This is easily integrated to give

$$f(t) = f(0) \left[ 1 - \frac{t}{t_{\text{rr}}} \right]^{-3/8}, \quad (3.2)$$

where the radiation reaction time scale  $t_{\text{rr}}$  is given by

$$t_{\text{rr}} = \frac{5}{256} \frac{1}{\pi^{8/3} f(0)^{8/3} \mu M^{2/3}}. \quad (3.3)$$

We assume that the gravitational-wave strain  $h_i(t)$  detected by LISA due to a binary (labeled by  $i$ ) located at distance  $D_i$  from the solar system assumes the form

$$h_i(t) = A_0 \frac{\mathcal{M}_i}{D_i} [\pi \mathcal{M}_i f_i(t)]^{2/3} \cos \left[ \varphi_i + 2\pi \int_0^t f_i(t') dt' \right] \quad (3.4)$$

where  $\varphi_i$  is a random initial phase and  $A_0$  is an overall factor of order one (discussed below). Again,  $f_i(t) = f_i(0)(1 - t/t_{\text{rr},i})^{-3/8}$ , and henceforth we will refer to  $f_i(0)$  as simply  $f_i$ . The quantity  $\mathcal{M}_i$  is the binary chirp mass defined as

$$\mathcal{M}_i = M_i^{2/5} \mu_i^{3/5}, \quad (3.5)$$

where  $M_i$  and  $\mu_i$  are the (locally measured) total and reduced masses, respectively.

Equation (3.4) is valid as long as the binary is close enough so that cosmological effects can be neglected. This is certainly true for galactic white dwarf binaries. However this is generally not the case for EMRIs or MBHBs, which will typically be at cosmological distances. For cosmologically distant binaries, we instead have [9]:

$$h_i(t) = A_0 \frac{\mathcal{M}_i(z_i)}{D_L(z_i)} [\pi \mathcal{M}_i(z_i) f_i(t)]^{2/3} \cos \left[ \varphi_i + 2\pi \int_0^t f_i(t') dt' \right]. \quad (3.6)$$

The quantity  $\mathcal{M}_i(z_i)$  is the redshifted chirp mass defined as

$$\mathcal{M}_i(z_i) = (1 + z_i) M_i^{2/5} \mu_i^{3/5}. \quad (3.7)$$

The quantity  $D_L(z)$  is the standard luminosity distance at redshift  $z$  for a flat universe, namely

$$D_L(z) = \frac{(1+z)}{H_0} \int_0^z \frac{dz'}{[\Omega_m(1+z')^3 + \Omega_\Lambda]^{1/2}}. \quad (3.8)$$

In this paper we use the values  $\Omega_m = 0.3$  and  $\Omega_\Lambda = 0.7$ . (It turns out that we do not require a precise value for  $H_0$  for our analyses, since this factor just gets absorbed into a quantity representing the total number of EMRIs out to some maximum redshift.)

In reality the overall factor  $A_0$  depends on the four angles in the problem (the source's sky location and orientation), and is in fact time varying due to LISA's changing antenna pattern. However for this paper we neglect those dependencies—in effect approximating  $A_0$  by its rms value. We would not expect this approximation to greatly affect the overall shape of  $p(x)$ . Moreover, we expect that at large  $x$ ,  $p(x)$  is dominated by background sources that are close (small  $D_i$ ), rather than ones with particularly favorable orientations. Since it is primarily the tail of  $p(x)$  that determines the behavior of  $P_N(X)$  at large  $X$ , we do not expect this averaging over angles to greatly affect our conclusions.

We write the confusion noise strain  $c(t)$  as follows

$$c(t) = \sum_i h_i(t) = \sum_i A_i(t) \cos \left[ \varphi_i + 2\pi \int_0^t f_i(t') dt' \right], \quad (3.9)$$

where

$$A_i(t) = A_0 \frac{\mathcal{M}_i}{D_i} [\pi \mathcal{M}_i f_i(t)]^{2/3}. \quad (3.10)$$

For our purpose, the quantity of interest is the overlap  $X \equiv (c|h)$  of a given normalized template  $h$  (from a given class of sought-for sources) with the confusion noise  $c$ :

$$(c|h) = \sum_i (h_i|h), \quad (3.11)$$

where the inner product  $(h_i|h)$  is defined as

$$(h_i|h) = 2 \int_{-\infty}^{+\infty} \frac{\tilde{h}_i(f) \tilde{h}^*(f)}{S_n(|f|)} df, \quad (3.12)$$

where  $\tilde{h}_i(f)$  and  $\tilde{h}(f)$  are the Fourier transforms of  $h_i(t)$  and  $h(t)$ , respectively, and  $S_n(|f|)$  is the one-sided noise spectral density. In all searches we consider in this paper, the template will also be a Newtonian chirp of the form (3.4)

$$\begin{aligned} h(t) &= A(t) \cos \left[ \varphi + 2\pi \int_0^t f(t') dt' \right] \\ &= 2 \sqrt{\frac{6\mu M^{2/3}}{5\pi I}} [\pi f(0)(1 - t/t_{\text{rr}})^{-3/8}]^{2/3} \\ &\quad \times \cos \left[ \varphi + 2\pi \int_0^t f(t') dt' \right], \end{aligned} \quad (3.13)$$

where the normalization condition  $(h|h) = 1$  implies

$$I = \int_0^\infty \frac{f^{-7/3} df}{S_n(f)}. \quad (3.14)$$

Both  $h_i(t)$  and  $h(t)$  are instantaneously monochromatic signals with slowly varying frequencies. Consider the tracks  $f_i(t)$  and  $f(t)$  that their frequencies sweep out in the  $t$ - $f$  plane. The integral in Eq. (3.12) is dominated by the point where the two tracks cross. Using the stationary phase approximation, the integral can be approximated as [10]

$$\begin{aligned} (h_i|h) &= \frac{1}{S_n[f_i(t_i)]} A_i(t_i) A(t_i) |\delta \dot{f}_i(t_i)|^{-1/2} \\ &\quad \times \cos[\delta \Phi_i + \text{sgn}(\delta \dot{f}_i) \pi/4], \end{aligned} \quad (3.15)$$

where

$$\begin{aligned} \delta \dot{f}_i &= [\dot{f}(t_i) - \dot{f}_i(t_i)] \\ &= \frac{3}{8} \frac{f(t_i)^{11/3}}{\pi^{8/3}} \left[ \frac{1}{t_{\text{rr}} f^{8/3}(0)} - \frac{1}{t_{\text{rr},i} f_i^{8/3}(0)} \right] \end{aligned} \quad (3.16)$$

and

$$\delta\Phi_i = \varphi - \varphi_i + 2\pi \int_0^{t_i} [f(t') - f_i(t')] dt'. \quad (3.17)$$

The time  $t_i$  is the instant of time when the template and the  $i$ th binary cross in the time-frequency plane, i.e. when  $f(t_i) = f_i(t_i)$ . (If the template and the  $i$ th binary either do not cross in the time-frequency plane, or cross outside LISA's sensitivity band or outside the observation period, then we approximate their overlap by zero. In the applications below this is implemented by restricting the integration range over binary parameters.)

A GW background is essentially a distribution of unresolved signals. In the next two subsections we introduce model distributions for the GWDB and EMRI backgrounds, respectively.

## B. Binary parameters for galactic white dwarf binaries

Here we present our model distribution for the galactic WD binaries. For simplicity, we will assume that all WD binaries have the same chirp mass, for which we adopt the median value arising from recent population synthesis calculations:  $M_c = 0.25M_\odot$  [11]. (This is approximately the  $M_c$  for a binary composed of two  $0.3M_\odot$  WDs.) We further assume that the other binary parameters are drawn from the following distributions:

$$p(D_i)dD_i = \theta(D_i - D_{\min}(f_i))\theta(D_{\max} - D_i) \times \frac{2D_i}{D_{\max}^2 - D_{\min}^2(f_i)} dD_i, \quad (3.18a)$$

$$p(f_i)df_i = \theta(f_i - f_{\min})\theta(f_{\max} - f_i) \frac{8}{3f_{\min}} \left(\frac{f_i}{f_{\min}}\right)^{-11/3} df_i, \quad (3.18b)$$

$$p(\varphi_i)d\varphi_i = \frac{d\varphi_i}{2\pi}, \quad (3.18c)$$

where, in (3.18b),  $f_i$  is the initial (i.e., at the beginning of the data set) gravitational-wave frequency of the binary, and  $f_{\min}$  and  $f_{\max}$  represent some low- and high-frequency cutoffs for the population we are considering. For our applications, we shall generally take  $f_{\min} = 10^{-4}$  Hz and  $f_{\max} = 10^{-2}$  Hz. The scaling  $p(f_i) \propto f_i^{-11/3}$  just comes from the assumption that binaries are “born” at frequencies below  $f_{\min}$  and then evolve according to Eq. (3.1). (Basically, binaries evolve much faster at higher frequency, and so are correspondingly sparser there.)

The distance probability distribution  $p(D_i)$  assumes that all galactic WD binaries are uniformly distributed in a disk of radius  $D_{\max} = 10$  kpc, centered on our Solar System. Clearly this is inaccurate in two ways. First, the Solar System is not at the center of the Milky Way (nor is the Milky Way a uniform disk). However the nonuniformity and noncenteredness (around us) of the galactic disk clearly mostly affects the distribution of binaries more distant than a few kpc, and these are not among the strongest sources. The distant binaries do not strongly affect the “high- $\sigma$ ” tail of

$p(x)$ , which is what is crucial for determining the tail of  $P_N(X)$ . Therefore this aspect of the uniform-disk assumption should be fairly harmless. More problematic is that the disk model departs significantly from reality at distances less than the thickness of the disk, which is  $\sim 600$  pc. Below this distance, it would be better to approximate the WD distribution as spherical. However we shall not do this for the following reason. Assuming that the distribution is planar clearly overestimates the number of nearby WD binaries, which artificially amplifies the high- $\sigma$  tail of  $p(x_i)$  (since the closest background sources have the largest overlaps with any searched-for signal). At the end of our analysis, we shall find that, even with the uniform-disk distribution, the non-Gaussianity of the WD background is a negligible factor in searching for MBHBs or EMRIs. Had we correctly modified  $p(D_i)$  for  $D < 600$  pc, the non-Gaussianity would still be negligible—even more so. That is, while the uniform-disk assumption is hard to justify *a priori*, it is fully justified *a posteriori*.

Finally, we discuss the inner cutoff  $D_{\min}(f_i)$ . The justification for imposing an inner cutoff  $D_{\min}(f_i)$  is that within this distance any WD binary of frequency  $f_i$  would be so bright that it could immediately be found in the data and essentially subtracted out, *before* searching for other types of signals. So when we state results, “the GWDB background” is really short for “the GWDB background minus the very brightest, immediately identifiable GWDB sources.”

What is a reasonable value for  $D_{\min}(f_i)$ ? A straightforward calculation shows that the (sky-averaged) LISA signal-to-noise for a white dwarf binary at distance  $D$  is

$$\text{SNR} = 1.26 \left( \frac{\mathcal{M}}{0.25 M_\odot} \right)^{5/3} \left( \frac{1 \text{ kpc}}{D} \right) \left( \frac{T_{\text{obs}}}{3 \text{ yr}} \right)^{1/2} \times \left( \frac{f}{1 \text{ mHz}} \right)^{11/6}. \quad (3.19)$$

This is the combined SNR from LISA's A and E channels, assuming that the noise is dominated by WD confusion noise with (sky-averaged) spectral density  $S_n(f) = S f^{-7/3}$ , with  $S = 1.44 \times 10^{-44} \text{ Hz}^4/3$ . Let  $\rho_{\text{th}}$  the signal-to-noise threshold, such that GWDBs with  $\text{SNR} > \rho_{\text{th}}$  are immediately subtracted from the data (or otherwise accounted for) *before* searching for (high- $z$ , weaker) MBHBs or EMRIs. We shall take  $\rho_{\text{th}} = 50$  as a reasonable fiducial value.

$$D_{\min}(f_i) = 2.5 \times 10^{-2} \left( \frac{50}{\rho_{\text{th}}} \right) \left( \frac{\mathcal{M}}{0.25 M_\odot} \right)^{5/3} \left( \frac{T_{\text{obs}}}{3 \text{ yr}} \right)^{1/2} \times \left( \frac{f_i}{1 \text{ mHz}} \right)^{11/6} \text{ kpc}. \quad (3.20)$$

It is absolutely crucial that there be *some* such threshold. Since  $x_i \propto 1/D_i$ , if there were no threshold we would have  $p(x_i) \propto x_i^{-3}$  at large  $x_i$ , and therefore  $P_N(X)$  would fall off only as  $X^{-3}$  at large  $X$ . We believe the cutoff is physically reasonable, since there is no reason one cannot subtract off

the very bright sources before looking for weaker ones. (Of course, at the very end of the data analysis one will want to find the joint best fit for all sources, which will involve readjusting the parameters of all the sources, including the ones that were initially “subtracted out”.)

Note, however, that we are imagining removing *only* the very strongest GWDBs in the chosen band. Now, we expect that LISA data analysis will actually proceed in stages, and our GWDB model basically represents this background at a rather early stage in the analysis. At a later stage, we expect that it will be possible to identify and subtract out all GWDBs with frequencies above a few mHz. In principle we could certainly adjust the value of  $f_{\max}$  in Eq. (3.18b) for different stages in the data analysis, but for simplicity in this paper we just adopt one fixed value for  $f_{\max}$ .

Finally the distribution (3.18c) for  $p(\varphi_i)$  simply states that the initial orbital phase (and hence also the GW phase) of each binary is random and uniformly distributed. This uniform distribution in initial phase is what leads to  $p(x) = p(-x)$ ; i.e. negative values of  $\langle h|h_i \rangle$  are just as likely as positive ones.

### C. Distribution of binary parameters for EMRIs

We next turn to unresolved EMRIs as a source of confusion noise. There are three types of EMRIs, since the inspiraling compact object can be a WD, a neutron star

(NS), or a black hole (BH). BC estimated the spectral density of confusion noise from each of these populations. The estimates are uncertain by at least an order of magnitude, due to the uncertainty in EMRI capture rates. EMRI confusion noise could end up being comparable to LISA’s instrumental noise, and perhaps even comparable to GWDB confusion noise, for rates at the high end of the estimated range. The WDs and NSs would at first seem to represent a bigger confusion problem, since more than 90% of the GW signal from NS and WD EMRIs will come from unresolvable sources; i.e., the NS and WD signals mostly represent confusion noise. BHs are more massive and so give stronger, more readily resolved signals; perhaps only 30% of the GW signal from all BH EMRIs is unresolvable. Nevertheless, since mass segregation tends to concentrate the heavier BHs closer to the MBH, and since supernova kicks may effectively empty the inner few parsecs of NSs, our judgement is that, of the three source types, BHs are the most likely to lead to substantial confusion noise. For this reason, and for simplicity, we consider a population model composed entirely of BH EMRIs. Since the total signal from BH EMRIs will be dominated by events at cosmological distances, our model takes cosmological effects into account, and we include the effects of evolution in both the MBH mass and the event rate. We adopt the following distribution as our population model:

$$p(M_i)dM_i = \theta[M_i - 10^5(1+z_i)^{-0.6}M_\odot]\theta[10^7(1+z_i)^{-0.6}M_\odot - M_i]\frac{17}{24}(10^{119/24} - 10^{85/24})^{-1}\left[\frac{M_i}{(1+z_i)^{-0.6}M_\odot}\right]^{17/24}\frac{dM_i}{M_i}, \quad (3.21a)$$

$$p(\mu_i)d\mu_i = \theta[\mu_i - 5M_\odot]\theta[15M_\odot - \mu_i]\frac{1}{\ln 3}\frac{d\mu_i}{\mu_i}, \quad (3.21b)$$

$$p(z_i)dz_i = \theta(z_i - z_c)\theta(2 - z_i)\mathcal{N}(z_c)[H_0D_L(z_i)]^2\frac{(1+z_i)^{-1.4}}{\sqrt{0.3(1+z_i)^3 + 0.7}}dz_i, \quad (3.21c)$$

$$p(\varphi_i)d\varphi_i = \frac{d\varphi}{2\pi}, \quad (3.21d)$$

$$p(f_i)df_i = \theta(f_i - f_{\min})\frac{8}{3f_{\min}}\left(\frac{f_i}{f_{\min}}\right)^{-11/3}df_i. \quad (3.21e)$$

Here  $M_i$  and  $\mu_i$  are the locally measured masses of the MBH and stellar-mass BH, respectively. The mass probability distributions (3.21a) and (3.21b) are derived from the following considerations. The actual distribution  $p(\mu_i)$  is very poorly known, but seems centered on  $\mu_i \approx 10M_\odot$ , so we simply assume a flat distribution between 5 and  $15M_\odot$ . For  $p(M_i)$  we restrict attention to the MBHs that *today* have masses between  $10^5$  and  $10^7M_\odot$ , and we assume that their masses have been increasing in time like  $t^{1/2} \propto (1+z)^{-0.6}$ , as they continuously swallow gas and compact objects. The locally measured distribution of compact object masses is assumed independent of time. The dependence of probability distributions (3.21a) and (3.21b) on  $M$

and  $\mu$  respectively is obtained from assumptions on the scaling of merger rates with masses. Let  $N$  be the number of mergers with masses comprised between  $M$  and  $M + dM$  and  $\mu$  and  $\mu + d\mu$ . The rate  $R$  of mergers within this box of mass parameters is

$$R = \frac{dN}{df}\dot{f}. \quad (3.22)$$

Following Gair *et al.* [2], we take the rate  $R$  to be proportional to  $M^{3/8}$ . Since we are considering only a rather narrow range (a factor of 3) of masses for the inspiraling object (and since the distribution of stellar BH masses is poorly known), we approximate  $R$  as being independent of



$\mu$ . These assumptions lead to the following scaling relation

$$\frac{dN}{df} \propto \frac{R}{\dot{f}} \propto \frac{M^{3/8}}{f^{11/3} \mu M^{2/3}} = f^{-11/3} \mu^{-1} M^{-7/24}. \quad (3.23)$$

From the definition of  $N$  we immediately obtain  $p(M) \propto M^{-7/24}$  and  $p(\mu) \propto \mu^{-1}$ . Note incidentally that this also gives the probability distribution (3.21e), which was previously derived using the fact that the probability of finding a binary between frequency  $f$  and  $f + df$  is proportional to the fraction of the binary lifetime it spends around that frequency.

We restrict attention to sources at redshift  $z \leq 2$ , partly since the rates at higher redshift are even more highly uncertain, and partly since the summed contribution from the  $z > 2$  sources, all individually weak, clearly will be much more nearly Gaussian than the noise from the  $z < 2$  population.

The distribution (3.21c) for  $p(z_i)$  is then obtained directly from Eq. (10) of Ref. [10], assuming that the locally measured capture rate  $\dot{n}$  scales as  $\dot{n} \propto (1+z)^{0.6}$ , i.e. the capture rate decreases over time as  $t^{-1/2}$ . This decrease reflects the fact that the MBH first swallows the closest objects, and then it has to wait longer and longer for further compact objects to diffuse in [2]. As with the case of galactic white dwarf binaries, we impose a short-distance cutoff  $z_c$ , reflecting the fact that very nearby sources can be easily identified and taken out of the confusion noise. In this paper we adopt the nominal value  $z_c = 0.1$ . BH EMRIs closer than  $z_c$  would typically have matched-filter SNRs in excess of 300 [2], and so should be very easily identified early in the data analysis. The normalization constant  $\mathcal{N}(z_c)$  [defined in Eq. (3.21c)] associated with this choice is  $\mathcal{N}(0.1) = 1.03\,044$ .

#### IV. APPLICATIONS TO VARIOUS SEARCHES

In this section we apply the statistical tools of Sec. II to the cases of matched-filter searches for MBHBs or EMRIs buried in confusion noise. As emphasized in Sec. I, we focus on the weakest resolvable signals of these types. We want to assess the importance of the non-Gaussian tails of the SNR distribution in setting the appropriate detection thresholds for these searches.

##### A. MBHB search: confusion from galactic white dwarf binaries

We first consider the problem of searching for MBHB signals buried in the confusion noise from GWDBs. As a particular MBHB template signal chirps upwards in frequency, its track on the  $t$ - $f$  plane intersects the tracks of all the white dwarf binaries in the galaxy (up to the final merger frequency of the MBHB). The GWDBs have random parameters, with PDF given by Eqs. (3.18a)–(3.18c), so each GWDB contributes some amount  $x_i$  to the detection statistic.

Our first goal is to obtain the PDF  $p(x_i)$ , from which we will estimate  $P_N(X)$  using the Edgeworth expansion. Our parent variable  $x_i$  is an individual overlap given by

$$x_i = \frac{1}{S_n[f_i(t_i)]} A_i(t_i) A(t_i) |\delta \dot{f}_i(t_i)|^{-1/2} \times \cos[\delta \Phi_i + \text{sgn}(\delta \dot{f}_i) \pi/4] \quad (4.1)$$

An important point to notice here is that since our template is a MBHB, it is chirping much faster than any GWDB. Thus to a good approximation we have

$$|\delta \dot{f}_i(t_i)| = \dot{f}(t_i). \quad (4.2)$$

In fact, the white dwarf binaries are chirping so slowly that it is reasonable to assume that their frequencies are constant for our analysis.

Equation (2.18) requires the first three nontrivial cumulants of the parent variable  $x_i$ , which are

$$\kappa_2 = E[x_i^2] = \sigma_x^2 \quad (4.3a)$$

$$\kappa_3 = E[x_i^3] \quad (4.3b)$$

$$\kappa_4 = E[x_i^4] - 3\sigma_x^4 \quad (4.3c)$$

Clearly, since the random phase  $\delta \Phi_i$  is uniformly distributed between 0 and  $2\pi$ , all odd cumulants of  $x_i$  vanish and so  $\kappa_3 = 0$ . We now compute  $\kappa_2$  and  $\kappa_4$ . For the second cumulant, we have

$$\begin{aligned} \kappa_2 &= \int x_i^2 p(D_i) dD_i p(f_i) df_i p(\varphi_i) d\varphi_i \\ &= \int p(D_i) dD_i p(f_i) df_i \frac{A_i^2(t_i) A^2(t_i)}{2S_n^2[f_i(t_i)]} \frac{1}{\dot{f}(t_i)} \end{aligned} \quad (4.4)$$

To go further we need a specific expression for  $S_n(f)$ . For simplicity we will approximate  $S_n(f)$  in the band  $[f_{\min}, f_{\max}]$  (with  $f_{\min} \sim 10^{-4}$  Hz and  $f_{\max} \sim 10^{-2}$  Hz) by the spectral density of the GWDB background (which does indeed work well throughout most of this band), *including* the contribution from GWDBs with  $f > 3$  mHz (which should be resolvable at a later stage of the data analysis). LISA's instrumental noise rises steeply below and above this band, so we approximate  $1/S_n(f)$  as vanishing outside it. Following [1], we therefore approximate  $1/S_n(f)$  by

$$\frac{1}{S_n(f)} = \theta(f - f_{\min}) \theta(f_{\max} - f) \frac{f^{7/3}}{S}, \quad (4.5)$$

with  $S = 1.44 \times 10^{-44}$  Hz<sup>4/3</sup>. This also implies  $I = (f_{\max} - f_{\min})/S$  [cf. Eq. (3.14)]. Combining (3.10), (3.13), (3.20), (4.2), and (4.5), we obtain the following general formula for all even raw moments:

$$E[x_i^{2n}] = \frac{(2n-1)!!}{(2n)!!} \left[ \frac{A_0^2 \mathcal{M}_i^{10/3}}{4\pi I S^2} \right]^n \times \int_{f_{\min}}^{f_{\max}} p(f_i) f_i^{11n/3} df_i \int_{D_{\min}(f_i)}^{D_{\max}} p(D_i) D_i^{-2n} dD_i. \quad (4.6)$$

Note that the lower bound in the frequency integral assumes that the initial template frequency  $f_0$  lies below the LISA band lower bound  $f_{\min}$ . If one is interested in templates that begin inside the LISA band, then the lower bound on the frequency integral should be replaced by  $f_0$ .

In evaluating the distance integrals, we shall assume here that  $D_{\max} \gg D_{\min}(f_i)$  for  $f_{\min} < f_i < f_{\max}$ . We can then approximate  $p(D_i)$  by

$$p(D_i) \simeq \theta(D_i - D_{\min}) \theta(D_{\max} - D_i) \frac{2}{D_{\max}^2} D_i, \quad (4.7)$$

which simplifies results considerably.<sup>2</sup> The computation of the remaining integrals are straightforward and the results are

$$E[x_i^2] = \frac{1}{2} \left[ \frac{A_0^2 \mathcal{M}_i^{10/3}}{4\pi I S^2} \right] \left[ \frac{16f_{\min}^{8/3}}{3D_{\max}^2} \right] \left( f_{\max} \left\{ \ln \left[ \frac{D_{\max}}{D_{\min}(f_{\max})} \right] + \frac{11}{6} \right\} - f_{\min} \left\{ \ln \left[ \frac{D_{\max}}{D_{\min}(f_{\min})} \right] + \frac{11}{6} \right\} \right) \quad (4.8a)$$

$$E[x_i^{2n}] = \frac{(2n-1)!!}{(2n)!!} \left[ \frac{A_0^2 \mathcal{M}_i^{10/3}}{4\pi I S^2} \right]^n \left[ \frac{16f_{\min}^{8/3}}{3D_{\max}^2} \right] \times \left[ \frac{(f_{\max} - f_{\min})(1 \text{ mHz})^{11(n-1)/3}}{2(n-1)D_{\min}^{2(n-1)}(1 \text{ mHz})} \right] \quad \text{for } n \geq 2. \quad (4.8b)$$

We then obtain the following expression for the fourth cumulant

$$\kappa_4 = \sigma_x^4 \left\{ \frac{9\bar{f}_{\min}^{-8/3}(\bar{f}_{\max} - \bar{f}_{\min})}{64} \left[ \frac{D_{\max}}{D_{\min}(1 \text{ mHz})} \right]^2 \times \left( \bar{f}_{\max} \left\{ \ln \left[ \frac{D_{\max}}{D_{\min}(f_{\max})} \right] + \frac{11}{6} \right\} - \bar{f}_{\min} \left\{ \ln \left[ \frac{D_{\max}}{D_{\min}(f_{\min})} \right] + \frac{11}{6} \right\} \right)^{-2} - 3 \right\}, \quad (4.9)$$

where  $\bar{f} = f/(1 \text{ mHz})$ . Since  $\bar{f}_{\min} \sim 0.1$  and  $\Delta^2 \equiv D_{\max}^2/D_{\min}^2(1 \text{ mHz}) = (10\rho_{\text{th}}/1.26)^2 \sim 10^5$ , the first term in (4.9) is clearly much larger than the second and so we can safely drop the  $-3$  term. Substituting the resulting fourth cumulant into (2.18) finally yields the following PDF for the signal-to-noise ratio

$$P_N(X) = \frac{1}{\sqrt{2\pi N \sigma_x^2}} e^{-(X^2/2N\sigma_x^2)} \left[ 1 + \frac{3\bar{f}_{\min}^{-8/3}(\bar{f}_{\max} - \bar{f}_{\min})\Delta^2}{512N} \times \left\{ (\bar{f}_{\max} - \bar{f}_{\min}) \ln \Delta + \frac{11}{6} \left[ \bar{f}_{\max}(1 + \ln \bar{f}_{\max}) - \bar{f}_{\min}(1 + \ln \bar{f}_{\min}) \right] \right\}^{-2} H_4 \left( \frac{X}{\sqrt{N\sigma_x^2}} \right) \right] \quad (4.10)$$

In the above the number  $N$  is the number of galactic white dwarf binaries with gravitational-wave frequency above  $f_{\min}$ , our assumed lower bound of the LISA band.<sup>3</sup> Using  $\bar{f}_{\min} \sim 0.1$ ,  $\bar{f}_{\max} \sim 10$ , a threshold  $\rho_{\text{th}} \sim 50$ , 3 years of observation and assuming  $N = 3 \times 10^7$  galactic binaries contributing to the confusion noise in the LISA band, the relative size of the  $1/N$  correction to the Gaussian PDF predicted by the central limit theorem at the  $7\sigma$  is  $\sim 0.02$ . Thus, at the  $7\sigma$  level, the Edgeworth analysis shows that the non-Gaussian tails of the SNR distribution is negligible for this search.

Finally, notice that if we take the inner cutoff distance  $D_{\min}(f)$  as fixed, then the result (4.10) is actually independent of our overall factor  $A_0$  in the waveform amplitudes, and also independent of the magnitude of  $S$  in Eq. (4.5) for the noise spectral density appearing in the inner product. Multiplying either of these by an overall factor simply rescales all the  $x_i$ ,  $X$ , and  $\sigma_x$  by the same amount, while the result (4.10) is expressed purely in terms of the dimensionless ratio  $X/\sigma_x$ .

## B. EMRI search: confusion from galactic white dwarf binaries

Notice here the important fact that our results in Sec. IVA are independent of the chirp mass  $M_c$  of the normalized search template. The reason for this is that while the template amplitude (for normalized templates) scales like  $M_c^{5/6}$  [cf. Eq. (3.13)], the time over which there is significant overlap with any GWDB signal scales like  $(\dot{f})^{-1/2} \propto M_c^{-5/6}$ . Therefore the analysis in Sec. IVA applies with practically no modification to searches for EMRIs buried in galactic white dwarf confusion noise. In the case of a realistic EMRI search in Gaussian noise, the detection threshold is around  $\sim 14\sigma$ . At that level the relative correction to  $P_N(X)$  predicted by the  $1/N$  term in the Edgeworth series is  $\sim 0.4$ . Since this correction is of order unity, one should also check the  $1/N^2$  term.

Performing that calculation using the results of the previous subsection and the next-to-leading term of Edgeworth series (given by, e.g., Petrov [5]) yields a next-to-leading correction of order  $\sim 0.2$ . Since this correction term is smaller than the leading term, we are

<sup>2</sup>For a signal-to-noise threshold  $\rho_{\text{th}} \sim 50$  and three years of observation, the ratio  $D_{\min}^2/D_{\max}^2$  is of order  $10^{-9}$  at 0.1 mHz,  $\sim 10^{-5}$  at 1 mHz and  $\sim 0.03$  at 10 mHz. Thus the error from this approximation is negligible throughout the band 0.1–10 mHz.

<sup>3</sup>Again, if one is interested in a template which begins inside the LISA band at  $t = 0$ , then the number  $N$  appearing in (4.10) is the number of galactic white dwarf binaries in the frequency interval  $(f_0, f_{\max})$ , with  $f_0$  being the initial template frequency.

inclined to trust the leading-order correction to the Gaussian distribution. This confirms that the Gaussian approximation is still reasonably accurate at the  $14\sigma$  level, for EMRIs buried in galactic white dwarf confusion noise.

### C. MBHB search: confusion from EMRIs

In this subsection we compute the signal-to-noise PDF for a matched-filter search for MBHB signals buried in confusion noise from unresolved EMRIs. At any given time it is expected that  $\sim 10^{4-6}$  unresolved EMRIs signals will be radiating GWs into the band 0.1–10 mHz, representing a significant source of confusion noise. When searching for MBHBs in a background of EMRIs, we may again assume that the parameter  $\delta\dot{f}_i$  appearing in stationary phase overlap (3.15) is entirely dominated by

the chirping MBHB. However, since the radiation reaction time scale for EMRIs is comparable to the LISA mission lifetime, our calculation must take into account that some EMRIs that are “live” (i.e., are premerger) at the beginning of LISA’s observation period will “die” (merge) before their  $t$ - $f$  track can be crossed by the MBHB’s  $t$ - $f$  track.

#### 1. Edgeworth expansion

As before, we wish to compute the cumulants of the parent distribution  $p(x_i)$  to obtain the Edgeworth expansion of  $P_N(X)$ . Following Sec. IVA, we first derive an expression for the raw moments of the parent distribution, i.e.

$$\begin{aligned} E[x^{2n}] &= \frac{(2n-1)!!}{(2n)!!} \int p(z_i) dz_i p(f_i) df_i p(M_i) dM_i p(\mu_i) d\mu_i \frac{A_i^{2n}(t_i) A^{2n}(t_i)}{S_n^{2n}[f(t_i)]} \frac{1}{|\dot{f}(t_i)|^n} \\ &= \frac{(2n-1)!!}{(2n)!!} \left[ \frac{A_0^2}{4\pi IS^2} \right]^n \int \frac{(1+z)^{10n/3}}{D_L^{2n}(z_i)} p(z_i) dz_i \int p(M_i) dM_i p(\mu_i) d\mu_i p(f_i) df_i M_i^{4n/3} \mu_i^{2n} f(t_i)^{11n/3} \\ &\quad \times \theta[f(t_i) - f_{\min}] \theta[f_{\max} - f(t_i)] \theta(t_i). \end{aligned} \quad (4.11)$$

The step function  $\theta(t_i)$  sets to zero the contribution from EMRIs whose  $t$ - $f$  tracks do not cross that of the MBHB template within the band  $[f_{\min}, f_{\max}]$  (since if  $t_i < 0$ , the tracks must cross at some frequency below  $f_{\min}$ ). In terms of the following parameters

$$\alpha_i = \left( \frac{f_i}{f_0} \right)^{8/3}, \quad (4.12a)$$

$$\beta_i = \frac{\mu M_i^{2/3}}{\mu_i M_i^{2/3}}, \quad (4.12b)$$

the crossing time  $t_i$  is given by

$$t_i = \frac{t_{\text{rr}}}{\alpha_i} \left( \frac{\alpha_i - 1}{1 - \beta_i^{-1}} \right), \quad (4.13)$$

which then gives

$$1 - \frac{t_i}{t_{\text{rr}}} = \frac{1}{\alpha_i} \left( \frac{\beta_i - \alpha_i}{\beta_i - 1} \right). \quad (4.14)$$

As before, we shall assume that  $f_0 < f_{\min}$ , i.e. the MBHB template begins outside the LISA band, but still at high-enough frequency that it has time to sweep through the LISA band during the mission lifetime. Then  $\alpha_i > 1$ , and combining this with the condition  $t_i \geq 0$  implies that  $\beta_i > \alpha_i$ .

Switching integration variables from  $(f_i, \mu_i)$  at fixed  $M_i$  to  $(\alpha_i, \beta_i)$ , we then obtain

$$\begin{aligned} E[x^{2n}] &= \frac{(2n-1)!!}{(2n)!!} \left[ \frac{A_0^2 \mu^2 M^{4/3} f^{11/3}}{4\pi IS^2} \right]^n \\ &\quad \times \int \frac{(1+z)^{10n/3}}{D_L^{2n}(z_i)} p(z_i) dz_i \int p(M_i) dM_i \\ &\quad \times \int_{\beta_{\min}}^{\beta_{\max}} \frac{d\beta_i}{\beta_i \ln 3} \int_{\alpha_{\min}}^{\infty} \alpha_{\min} \frac{d\alpha_i}{\alpha_i^2} \beta_i^{-2n} \alpha_i^{11n/8} \\ &\quad \times \left( \frac{\beta_i - \alpha_i}{\beta_i - 1} \right)^{-11n/8} \theta[f(t_i) - f_{\min}] \\ &\quad \times \theta[f_{\max} - f(t_i)] \theta(\beta_i - \alpha_i), \end{aligned} \quad (4.15)$$

where  $\alpha_{\min} \simeq (f_{\min}/f_0)^{8/3}$ , the exact expression given below in (4.17).

Using (4.14), we can rewrite the step functions as follows

$$\theta[f(t_i) - f_{\min, \max}] = \theta[\alpha_i - \alpha_{\min, \max}], \quad (4.16)$$

where

$$\alpha_{\min, \max} = \left( \frac{f_{\min, \max}}{f_0} \right)^{8/3} \left[ 1 - \frac{1}{\beta_i} + \frac{1}{\beta_i} \left( \frac{f_{\min, \max}}{f_0} \right)^{8/3} \right]^{-1}. \quad (4.17)$$

We then get

$$\begin{aligned}
E[x^{2n}] &= \frac{(2n-1)!!}{(2n)!!} \left[ \frac{A_0^2 \mu^2 M^{4/3} f^{11/3}}{4\pi I S^2} \right]^n \\
&\times \int \frac{(1+z)^{10n/3}}{D_L^{2n}(z_i)} p(z_i) dz_i \int p(M_i) dM_i \\
&\times \int_{\beta_{\min}}^{\beta_{\max}} \frac{d\beta_i}{\beta_i \ln 3} \int_{\alpha_{\min}}^{\alpha_{\max}} \alpha_{\min} \frac{d\alpha_i}{\alpha_i^2} \beta_i^{-2n} \\
&\times \alpha_i^{11n/8} \left( \frac{\beta_i - \alpha_i}{\beta_i - 1} \right)^{-11n/8}. \quad (4.18)
\end{aligned}$$

Strictly speaking the upper integration limit over  $\alpha_i$  should be  $\min(\alpha_{\max}, \beta_i)$ . However we show below in (4.20) that  $\beta_i > \alpha_{\max}$ , which justifies our limit in the previous equation. Performing the  $\alpha_i$  integral yields

$$\begin{aligned}
E[x^{2n}] &= \frac{(2n-1)!!}{(2n)!!} \left[ \frac{A_0^2 \mu^2 M^{4/3} f^{11/3}}{4\pi I S^2} \right]^n \frac{\alpha_{\text{th}}}{\ln 3} \\
&\times \int \frac{(1+z)^{10n/3}}{D_L^{2n}(z_i)} p(z_i) dz_i \int p(M_i) dM_i \\
&\times \int_{\beta_{\min}}^{\beta_{\max}} \frac{8(\beta_i - 1)^{11n/8}}{(11n - 8)\beta_i^2} \\
&\times [\alpha_i^{11n/8-1} (\beta_i - \alpha_i)^{-11n/8+1}]_{\alpha_{\min}}^{\alpha_{\max}} d\beta_i. \quad (4.19)
\end{aligned}$$

Now from (4.17), we have

$$\beta_i - \alpha_{\min, \max} = \left( \frac{f_{\min, \max}}{f_0} \right)^{-8/3} (\beta_i - 1) \alpha_{\min, \max}, \quad (4.20)$$

which, when substituted into (4.19), yields the following

$$\begin{aligned}
E[x^{2n}] &= \frac{(2n-1)!!}{(2n)!!} \left[ \frac{A_0^2 \mu^2 M^{4/3} f^{11/3}}{4\pi I S^2} \right]^n \frac{\alpha_{\text{th}}}{\ln 3} \\
&\times \int \frac{(1+z)^{10n/3}}{D_L^{2n}(z_i)} p(z_i) dz_i \int p(M_i) dM_i \\
&\times \int_{\beta_{\min}}^{\beta_{\max}} \frac{8(\beta_i - 1)}{(11n - 8)\beta_i^2} \left[ \left( \frac{f_{\max}}{f_0} \right)^{(11n-8)/3} \right. \\
&\left. - \left( \frac{f_{\min}}{f_0} \right)^{(11n-8)/3} \right] d\beta_i. \quad (4.21)
\end{aligned}$$

Now since  $\beta_i \sim 10^5$  over the integration range  $(\beta_{\min}, \beta_{\max})$ , we may set  $\beta_i - 1 \simeq \beta_i$  to very good accuracy and perform the remaining  $\beta_i$  integral, which simply gives  $\ln 3$ , the normalization constant of the  $\beta_i$  probability distribution. Then the  $M_i$  integral trivially gives unity, as the remaining integrand, apart from  $p(M_i)$ , is independent of  $M_i$ . We are then left with

$$\begin{aligned}
E[x^{2n}] &= \frac{(2n-1)!!}{(2n)!!} \frac{8\alpha_{\text{th}}}{(11n-8)} \left[ \frac{A_0^2 \mu^2 M^{4/3} f^{11/3}}{4\pi I S^2} \right]^n \\
&\times \left\{ \left[ \frac{f_{\max}}{f_0} \right]^{(11n-8)/3} - \left[ \frac{f_{\min}}{f_0} \right]^{(11n-8)/3} \right\} \\
&\times \int \frac{(1+z)^{10n/3}}{D_L^{2n}(z_i)} p(z_i) dz_i. \quad (4.22)
\end{aligned}$$

The redshift integrals must be performed numerically and we denote each value as

$$\int_{z_{\min}}^{z_{\max}} \frac{(1+z)^{10n/3}}{D_L^{2n}(z_i)} p(z_i) dz_i \equiv H_0^{2n} \zeta_n. \quad (4.23)$$

For inner cutoff redshift  $z_{\min} = 0.1$  and  $z_{\max} = 2$ , the first few values of  $\zeta_n$  are

$$\zeta_1 = 4.179\,24 \quad (4.24a)$$

$$\zeta_2 = 29.722\,32 \quad (4.24b)$$

$$\zeta_3 = 763.5828. \quad (4.24c)$$

The first two terms of the Edgeworth-expanded PDF for the signal-to-noise ratio are then given by

$$\begin{aligned}
P_N(X) &= \frac{1}{\sqrt{2\pi N \sigma_x^2}} e^{-(X^2/2N\sigma_x^2)} \left\{ 1 + \frac{9}{1792N} \right. \\
&\times \left[ \frac{(1-x^{14/3})}{x^{8/3}(1-x)^2} \frac{\zeta_2}{\zeta_1^2} \right] H_4\left(\frac{X}{\sqrt{N\sigma_x^2}}\right) + O(N^{-2}) \Big\}, \quad (4.25)
\end{aligned}$$

where  $N$  is the number of EMRIs that lie in the LISA band and where  $x \equiv f_{\min}/f_{\max}$ . Taking  $N = 5 \times 10^5$ , the same frequency limits for the LISA band as before and using (4.24), the relative size of the  $1/N$  correction to the Gaussian PDF predicted by the central limit theorem at the  $7\sigma$  level (i.e., at  $X/\sqrt{N\sigma_x^2} = 7$ ) is found to be  $\approx 8$ . Since this “first-order correction” is already larger than the zeroth-order estimate, the Edgeworth expansion simply cannot provide a reliable answer for this problem. Instead, the problem of searching for MBHBs buried in EMRI confusion noise must be addressed within the context of the theory of large deviations, to which we turn next.

## 2. Large-deviations analysis

Here we compute the signal-to-noise PDF for a search or MBHBs buried in EMRI confusion noise, following the prescription of large-deviations theory. The starting point is the construction of the modified cumulant generating functional  $\lambda(\beta)$ . Since we have already computed analytically all the raw moments of the parent distribution, we may evaluate  $\lambda(\beta)$  from its power series expansion numerically to any desired accuracy, i.e. we use

$$\begin{aligned}
e^{\lambda(\beta)} &= \sum_{p=0}^{\infty} \frac{\beta^p}{p!} E[x^p] \\
&= 1 + \frac{1}{2}(\beta\sigma_x)^2 + \sum_{n=2}^{\infty} \frac{(\beta\sigma_x)^{2n}}{(2n)!} \frac{E[x^{2n}]}{\sigma_x^{2n}} \quad (4.26)
\end{aligned}$$

Defining  $\tilde{\beta} = \beta\sigma_x$  and using (4.22), we obtain the following expression for the cumulant generating functional



$$e^{\lambda(\beta)} = 1 + \frac{1}{2}\tilde{\beta}^2 + 8x^{8/3} \sum_{n=2}^{\infty} \frac{\tilde{\beta}^{2n}}{n!^2} \left[ \frac{1 - x^{(11n-8)/3}}{11n-8} \right] \zeta_n, \quad (4.27)$$

where

$$\tilde{\beta} = \left[ \frac{3}{16x^{8/3}(1-x)\zeta_1} \right]^{1/2} \tilde{\beta}. \quad (4.28)$$

We then compute the rate function describing the signal-to-noise PDF following these steps. First we compute the cumulant generating functional  $\lambda$  from (4.27) numerically. The infinite sum is truncated when the  $n$ th term of the sum is of order  $10^{-10}$  of the sum of the previous  $n-1$  terms. Because the sum converges<sup>4</sup> for any value of  $\beta$ , we are confident that this is a reasonable accuracy criterion. We next compute  $\lambda' \equiv d\lambda/d\tilde{\beta}$  (here  $\lambda$  is considered an implicit function of  $\tilde{\beta}$ ) by taking a derivative of (4.27) and evaluating the sum numerically using the same truncation criterion as before. The next step is the maximization over  $\beta$  (or equivalently  $\tilde{\beta}$ ) of the quantity  $I = Z\beta - \lambda = \tilde{Z}\tilde{\beta} - \lambda$ , where  $\tilde{Z} = Z/\sigma_x$ . The value of  $\tilde{\beta}$  which maximizes  $I$  is simply the one satisfying  $\tilde{Z} = \lambda'(\tilde{\beta})$ . By inverting numerically the function  $\lambda'$ , we obtain the function  $\tilde{\beta}(\tilde{Z})$ . The rate function can then be computed numerically for any desired value of  $\tilde{Z}$  as follows

$$I(\tilde{Z}) = \tilde{Z}\tilde{\beta}(\tilde{Z}) - \lambda[\tilde{\beta}(\tilde{Z})]. \quad (4.29)$$

The resulting rate function of the signal-to-noise PDF for the MBHB search in EMRI confusion noise is plotted in Fig. 1. The vertical axis is the actual rate function normalized by the central limit theorem estimate:  $I(Z)/[-0.5Z^2/\sigma_x^2]$ . The horizontal axis is the SNR normalized to unit standard deviation, or  $N^{1/2}Z/\sigma_x$ . The result is plotted for  $N = 5 \times 10^5$  in-band EMRIs, but to obtain the curve for any other value of  $N$ , one simply rescales the  $x$  labels on the horizontal axis by  $\sqrt{N/(5 \times 10^5)}$ .

From that figure, one can easily see that at the  $7\sigma$  level, the rate function derived from large-deviations theory differs significantly from the central limit theorem estimate (confirming our conclusion from the Edgeworth analysis in Sec. IV C 1). While the central limit theorem estimate for  $P_N(X = 7N^{1/2}\sigma_x)$  is  $(2\pi)^{-1/2} \exp[-49/2]$ , the actual probability density is  $\approx \exp[-0.72 \times 49/2]$ , or a factor  $\sim 10^3$  larger. Therefore in deciding the appropriate detection threshold, one must take into account the non-Gaussianity of the signal-to-noise PDF. In the next subsection we discuss the proper adjustment of the detection threshold, based on the rate function of Fig. 1.

<sup>4</sup>This is a consequence of the fact that the parent probability distribution has compact support in our model.

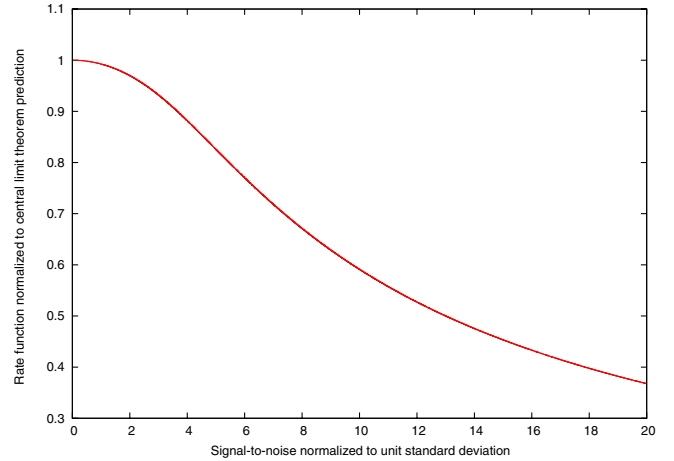


FIG. 1 (color online). This figure shows the rate function of the signal-to-noise PDF for the MBHB search in EMRI confusion noise for a simple Newtonian chirp toy model. The rate function is normalized to the value predicted by the central limit theorem, namely  $I_{\text{CLT}} = \tilde{Z}^2/2$ . This plot was generated using  $N = 5 \times 10^5$  for the number of unresolved EMRIs in the band 0.1–10 mHz. If one wants to vary the number of unresolved EMRIs, one simply rescales  $x$  values by  $\sqrt{N/(5 \times 10^5)}$ , since this is the multiplicative factor needed to rescale  $\tilde{Z}$  to unit standard deviation.

#### D. Adjusting the detection threshold

We have considered searches for MBHB signals buried in two different types of confusion background: GWDBs and EMRIs. For GWDBs we showed that the PDF for the SNR could be safely approximated as Gaussian (up to the detection threshold), but that a search in EMRI confusion noise alone would have to take into account the significant non-Gaussianity in  $P_N(X)$  at  $X \sim 7\sigma$ . However, EMRI confusion noise is unlikely to dominate the total noise, so in this subsection we show explicitly how to calculate the appropriate detection threshold for noise that is a sum of EMRI confusion noise plus Gaussian noise.

Since confusion noise from GWDBs is Gaussian to a good approximation, it can be combined with instrumental noise into one single source of Gaussian noise. We shall here denote the signal-to-noise ratio obtained by correlating a MBHB template with this Gaussian noise as  $\rho_g$ . This signal-to-noise is drawn from the following PDF

$$p_g(\rho_g) = \frac{1}{(2\pi\sigma_g^2)^{1/2}} e^{-\rho_g^2/2\sigma_g^2}, \quad (4.30)$$

where  $\sigma_g$  is the standard deviation of the random variable  $\rho_g$ . Next we denote the signal-to-noise ratio obtained by correlating a MBHB template with EMRI confusion noise as  $\rho_c$ . This signal-to-noise is drawn from the following PDF

$$p_c(\rho_c) = \mathcal{N}_c \exp\left[-\frac{\rho_c^2}{2\sigma_c^2} \tilde{I}\left(\sqrt{\frac{N^*}{N}} \frac{\rho_c}{\sigma_c}\right)\right], \quad (4.31)$$

where  $\tilde{I}$  is the rescaled rate function plotted in Fig. 1,  $N^* = 5 \times 10^5$  is the number of unresolved EMRIs chosen to generate Fig. 1,  $\mathcal{N}_c$  is a normalization constant and  $\sigma_c$  is the standard deviation of the random variable  $\rho_c$ . Consider now the PDF for the total signal-to-noise ratio  $\rho = \rho_g + \rho_c$ . It is given by the following convolution integral

$$p(\rho) = \int_{-\infty}^{+\infty} p_g(\rho - \rho_c) p_c(\rho_c) d\rho_c, \quad (4.32)$$

which can be performed numerically. If the rate function  $\tilde{I}$  were equal to unity, i.e. if  $\rho_c$  were Gaussian, then  $\rho$  would also be a Gaussian random variable with standard deviation  $\sigma = (\sigma_g^2 + \sigma_c^2)^{1/2}$ . We shall determine a threshold on the normalized total SNR  $\hat{\rho} \equiv (\sigma_g^2 + \sigma_c^2)^{-1/2} \rho$ , assuming that if the rate function  $\tilde{I}$  were equal to unity, then the appropriate detection threshold would be set at  $\hat{\rho} = 7$ . In other words, the acceptable false alarm probability  $P_{FA}$  is assumed to be the integral of the Gaussian PDF for  $\hat{\rho}$  over the range  $(-\infty, -7)$  and  $(7, +\infty)$ :

$$P_{FA} = 2 \int_7^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\hat{\rho}^2/2} d\hat{\rho} = \text{erfc}(7/\sqrt{2}). \quad (4.33)$$

The actual detection threshold  $\hat{\rho}_{th}$  for the MBHB search is then determined by the following equation

$$\begin{aligned} P_{FA} &= 2\sigma \int_{\hat{\rho}_{th}}^{\infty} p(\sigma\hat{\rho}) d\hat{\rho} \\ &= 2\sigma \int_{\hat{\rho}_{th}}^{\infty} \int_{-\infty}^{+\infty} p_g(\sigma\hat{\rho} - \rho_c) p_c(\rho_c) d\rho_c d\hat{\rho} \\ &= \int_{-\infty}^{+\infty} \text{erfc}\left[\frac{(1 + \varepsilon^2)^{1/2} \hat{\rho}_{th} - \varepsilon \hat{\rho}_c}{\sqrt{2}}\right] p_c(\rho_c) d\rho_c, \end{aligned} \quad (4.34)$$

where  $\hat{\rho}_c \equiv \rho_c/\sigma_c$  and where  $\varepsilon \equiv \sigma_c/\sigma_g$  measures the relative strength of the non-Gaussian component of the noise.

In Fig. 2 we plot  $\hat{\rho}_{th}/7$  as function of  $\varepsilon$  for our best estimate of  $\sigma_g$  (taken from BC, assuming no GWDBs have been fitted out). That is, we fix the amplitude of the GWDB background and plot how  $\hat{\rho}_{th}$  varies as one increases the number of unresolved in-band EMRIs. This figure was generated as follows. For any  $\varepsilon$  we estimated  $N$  (the number of unresolved EMRIs) using

$$\varepsilon \approx \left(\frac{N}{1.25 \times 10^7}\right)^{1/2}. \quad (4.35)$$

(Since an astrophysically reasonable estimate is  $N = 5 \times 10^5$ , we expect  $\varepsilon \approx 0.2$  in practice.) We insert  $N$  into Eq. (4.31) to obtain  $p_c(\rho_c)$ , which we then plug into the

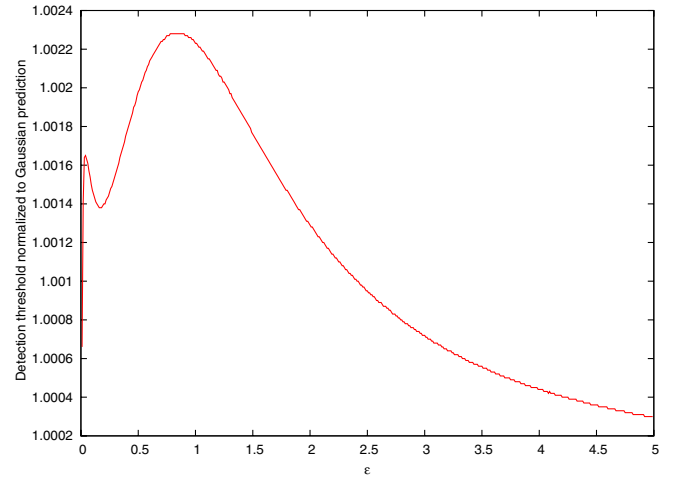


FIG. 2 (color online). This figure shows the normalized detection threshold  $\hat{\rho}_{th}$  (at fixed false alarm probability) for a total noise composed of a Gaussian component (instrumental noise and GWDB confusion noise) and a non-Gaussian component (EMRI confusion noise) as a function of the ratio  $\varepsilon = \sigma_c/\sigma_g$ . In this plot we consider the Gaussian component to be fixed and  $\varepsilon$  varies by adjusting the number of unresolved EMRIs. Note that  $\hat{\rho}_{th}$  is always nearly one, i.e., nearly the same as for a Gaussian distribution with the same standard deviation.

last line of Eq. (4.34). We obtain the detection threshold  $\hat{\rho}_{th}$  by solving (4.34) numerically.

The most important fact one gleans from Fig. 2 is that  $\hat{\rho}_{th}$  is always very close to 1. We can understand this as follows. For realistic values of  $N$ , the SNR from EMRIs is significantly non-Gaussian, but since the noise is dominated by instrumental and GWDB background noise, the non-Gaussianity of the EMRI background has little effect

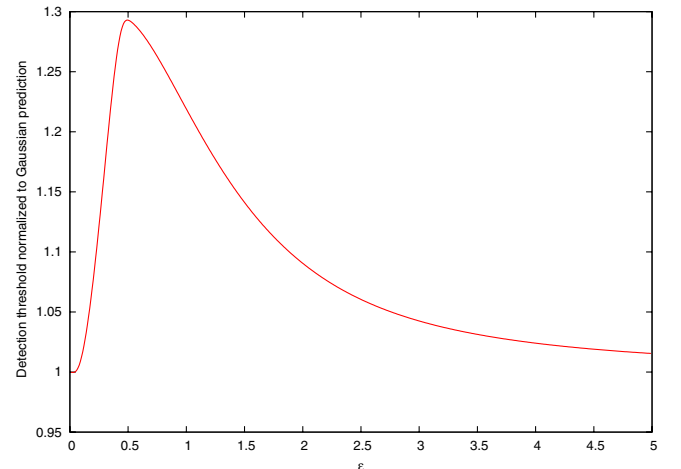


FIG. 3 (color online). This plot is similar to Fig. 2, but with the strength of the Gaussian component reduced “by hand” by a factor of 50, i.e.  $\sigma_g^2 \rightarrow \sigma_g^2/50$  so that  $\varepsilon = \sqrt{N}/(2.5 \times 10^5)$ . In this case, the detection threshold can be up to  $\sim 30\%$  higher than for a Gaussian distribution with the same standard deviation.

on the threshold. When  $N$  is large enough that EMRI noise is a large fraction of the total noise, the EMRI confusion noise is much more Gaussian, so again the threshold is very close to the Gaussian prediction.

Figure 3 is the same as Fig. 2, except that for illustrative purposes we have decreased “by hand” the value of  $\sigma_g$  by  $\sqrt{50}$ . In this case, the normalized threshold  $\hat{\rho}_{\text{th}}$  could be (for  $\epsilon \approx 0.5$ ) up to  $\sim 1.3$  times higher than for Gaussian noise with the same standard deviation. For  $N = 5 \times 10^5$  and this reduced  $\sigma_g$ , we would have  $\epsilon \approx 1.4$  and  $\hat{\rho}_{\text{th}} \approx 1.15$ ; i.e., the appropriate threshold would be  $8\sigma$  instead of  $7\sigma$ . We note that  $(\hat{\rho}_{\text{th}}/7) \rightarrow 1$  both as  $\epsilon \rightarrow 0$  and as  $\epsilon \rightarrow \infty$ . This is easily understood, since as  $\epsilon \rightarrow 0$  the noise becomes just the Gaussian part, while as  $\epsilon \rightarrow \infty$  we also have  $N \rightarrow \infty$ , so the EMRI portion becomes Gaussian.

## V. SUMMARY, CONCLUSIONS, AND OPEN ISSUES

In this paper we have analyzed the problem of determining the appropriate detection threshold for several idealized searches. The most important simplifications were that we used “lowest-order” waveforms (based on the quadrupole formula, and assuming quasicircular inspirals) and simplified population distributions for the confusion sources. We first considered searches for both MBHB signals and EMRI signals buried in confusion noise from GWDBs. Using the Edgeworth expansion, we showed that for these cases the PDF of the standard detection statistic remains nearly Gaussian out to the relevant detection thresholds. We then considered searches for MBHB signals buried in just EMRI confusion noise. In that case, using large-deviations theory, we found that  $7\sigma$  events would occur  $10^3$  times more often than suggested by the central limit theorem. However this third case was rather unrealistic, since it is very unlikely that EMRI confusion noise will dominate the total LISA noise curve. We then considered a more realistic example, in which the EMRI confusion noise was combined with Gaussian noise of  $\sim 5$  times larger amplitude. In that case, we again found that the non-Gaussianity of the EMRI confusion noise ends up having a negligible impact in setting the appropriate detection threshold.

The rather minimal impact of non-Gaussian tails in these models appears to stem from three circumstances. First, the number of confusion noise sources is always rather large. Second, in all cases we imposed a short-distance cutoff on the distribution of the background sources, arguing that the very closest and therefore strongest of the background sources could be effectively removed (or otherwise taken into account) *before* searching for other types of sources. Third, all three model problems shared the feature that the search templates and background templates evolve in frequency on very different time scales:  $\dot{f}_{\text{WD}} \ll \dot{f}_{\text{EMRI}} \ll \dot{f}_{\text{MBHB}}$ . Since  $x_i \propto |\delta \dot{f}_i|^{-1/2}$ , this separation of time scales ensures that  $p(x)$  has no large outliers arising from coincidentally small  $|\delta \dot{f}_i|^{-1/2}$ . Put another way, the dissimilar-

ity of the searched-for and background signals is crucial to the sharp falloff of  $p(x)$  at large  $x$ . The high- $X$  tail of  $P_N(X)$  depends crucially on the high- $x$  tail of  $p(x)$ , and the dissimilarity of the searched-for and background signals helps ensure a very steep falloff for  $p(x)$ .

We emphasize, however, that LISA data analysis will also present confusion noise problems where there is no such separation of time scales. For instance, consider the search for relatively nearby EMRIS signals embedded in the background noise from all the unresolvably distant EMRIs. In that case the parent distribution  $p(x_i)$  would surely have a substantial tail, due to cases where  $\delta \dot{f}_i$  is coincidentally small. Additionally, that detection problem raises issues of principle that we were not forced to confront in the model problems considered in this paper, and which we do not yet see how to resolve. For example, consider a case where some detection template  $\tilde{A}$  has overlap of 5, 4, 3, and 2 with background signals A, B, C, and D, respectively. Then the total SNR is 14 (assuming the sum of all other overlaps can be neglected), which naively might lead one to claim a detection. Should one consider that claim as a false alarm? What if most (but not all) of the parameters characterizing  $\tilde{A}$  are fairly close to those of A? Presumably experience with analyzing large sets of simulated data, as in the current mock LISA data challenges, will alert us if such issues arise very often in practice. However if such issues arise only rarely, then our experience with this project suggests that a sound theoretical understanding of the tails of the distribution could be crucial, since even with powerful computer clusters it could be difficult to sample the tails adequately with simulations.

## ACKNOWLEDGMENTS

É.R. was supported by NASA ATP Grant No. NNG04GK98G. C.C.’s work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract to the National Aeronautics and Space Administration.

## APPENDIX: HEURISTIC INTRODUCTION TO LARGE-DEVIATIONS THEORY AND CORRESPONDENCE WITH EDGEWORTH EXPANSION

Here we give a heuristic derivation of Chernoff’s formula in large-deviations theory and discuss its relation to the Edgeworth expansion. As before, let the sample mean be

$$Z = \frac{1}{N} \sum_{i=1}^N x_i. \quad (\text{A1})$$

Now consider the modified cumulant generating functional  $\Lambda(\tilde{\beta})$  for the sample mean PDF  $P_N(Z)$ . It is given by

$$\Lambda(\tilde{\beta}) = \ln \int e^{\tilde{\beta}Z} P_N(Z) dZ = N \ln \int e^{\tilde{\beta}x/N} p(x) dx$$

$$= N\lambda(\beta), \quad (\text{A2})$$

where  $\beta = \tilde{\beta}/N$  and where  $\lambda$  is the modified cumulant generating functional of the parent distribution. Now  $e^{\tilde{\beta}Z}$  is a rapidly increasing function of  $Z$ , while  $P_N(Z)$  is rapidly decreasing. Therefore one expects the integrand to be sharply peaked, and the integral to be some constant  $C$  of order 1 times that the value of the integrand at that maximum. [Of course, this is just Laplace's method of estimating the integral (A2).] Define  $S(Z) \equiv -\ln P_N(Z)$ . Then we have just argued that  $\Lambda(\tilde{\beta})$  is well approximated by

$$\Lambda(\tilde{\beta}) = \max_Z \{\tilde{\beta}Z - S(Z)\}. \quad (\text{A3})$$

That is,  $\Lambda(\tilde{\beta})$  is the Legendre transform of  $S(Z)$ , which we can invert to obtain

$$S(Z) = \max_{\tilde{\beta}} \{Z\tilde{\beta} - \Lambda(\tilde{\beta})\} \quad (\text{A4})$$

$$= N \max_{\beta} \{Z\beta - \lambda(\beta)\}. \quad (\text{A5})$$

Alternatively we may write

$$P_N(Z) = C e^{-NI(Z)} \quad (\text{A6})$$

where

$$I(Z) = \max_{\beta} [\beta Z - \lambda(\beta)] \quad (\text{A7})$$

and  $C$  is a normalization constant determined *a posteriori*. Clearly  $C$  is approximately given by  $C \approx [NI''(0)/(2\pi)]^{1/2}$ . This concludes our heuristic derivation of Chernoff's formula.

As a pedagogical example, consider the random variable  $X$  defined as

$$X = \sum_{i=1}^N x_i, \quad (\text{A8})$$

where  $x_i$  is a random variable equalling  $+1$  or  $-1$  with equal probability. The exact probability distribution for  $X$  is a binomial, i.e.

$$P_N(X) = \frac{N!}{\left(\frac{N+X}{2}\right)! \left(\frac{N-X}{2}\right)!} \left(\frac{1}{2}\right)^{N+1}$$

$$= \frac{N!}{\left[\frac{N}{2}\left(1 + \frac{Y}{\sqrt{N}}\right)\right]! \left[\frac{N}{2}\left(1 - \frac{Y}{\sqrt{N}}\right)\right]!} \left(\frac{1}{2}\right)^{N+1}, \quad (\text{A9})$$

where  $Y = X/\sqrt{N}$  measures how many standard deviations away from the mean the variable  $X$  lies. As  $N$  tends to infinity, we make use of the following refined version of Stirling's formula [12] to approximate (A9) as

$$n! = \sqrt{2\pi n} n^{n+1/2} e^{-n} \exp\left[\frac{1}{12n} - \frac{\theta_n}{360n^3}\right], \quad (\text{A10})$$

where the  $\theta_n$  are all bounded between 0 and 1. By substituting (A10) into (A9) and using  $Z = Y/\sqrt{N}$ , we obtain

$$P_N(Z) = \frac{1}{\sqrt{2\pi N}} (1+Z)^{-N(1+Z)/2-1/2} (1-Z)^{-N(1-Z)/2-1/2}$$

$$\times \exp\left[\frac{1}{12N} \left(1 - \frac{2}{(1+Z)} - \frac{2}{(1-Z)}\right) - \frac{1}{360N^3} \left(\theta_N - \frac{8\theta_{N(1+Z)/2}}{(1+Z)^3} - \frac{8\theta_{N(1-Z)/2}}{(1-Z)^3}\right)\right] \quad (\text{A11})$$

By further expanding in terms of  $Z \ll 1$  and keeping the leading-order corrections in  $1/N$ , we obtain

$$P_N(X) = \frac{1}{\sqrt{2\pi N}} e^{-(X^2/2N)} \exp\left[\frac{1}{N} \left(-\frac{1}{4} + \frac{X^2}{2N} - \frac{X^4}{12N^2}\right) \times [1 + O(N^{-1})]\right], \quad (\text{A12a})$$

$$P_N(Y) = \frac{1}{\sqrt{2\pi}} e^{(-Y^2/2)} \exp\left[\frac{1}{N} \left(-\frac{1}{4} + \frac{1}{2}Y^2 - \frac{1}{12}Y^4\right) \times [1 + O(N^{-1})]\right]. \quad (\text{A12b})$$

Let us now derive the large-deviations prediction for  $P_N(Z)$ . First the cumulant generating functional is given by

$$\lambda(\beta) = \ln E[e^{\beta x}] = \ln(\cosh \beta). \quad (\text{A13})$$

Maximizing  $I(Z)$  then yields

$$Z = \frac{d\lambda(\beta)}{d\beta} = \tanh \beta. \quad (\text{A14})$$

The rate function is therefore given by

$$I(Z) = Z \operatorname{arctanh} Z - \ln[\cosh(\operatorname{arctanh} Z)]$$

$$= \frac{1}{2} (1+Z) \ln(1+Z) + \frac{1}{2} (1-Z) \ln(1-Z). \quad (\text{A15})$$

This yields

$$P_N(Z) = C (1+Z)^{-N(1+Z)/2} (1-Z)^{-N(1-Z)/2} \quad (\text{A16})$$

for some normalization constant  $C$ . Comparing this with (A11), we see that it matches exactly the first line of (A11), neglecting the small  $-1/2$  term in each exponent. Large-deviations theory however does not capture the higher-order correction terms provided by that small  $-1/2$  term and the entire second line of (A11), which one needs to obtain expansion (A12a). How does large-deviations theory “fit in” with the central limit theorem and Edgeworth expansion? For simplicity let us assume that  $p(x)$  is an even function [i.e.,  $p(-x) = p(x)$ ]; clearly  $P_N(Z)$  is then also even. Presumably the exponent  $NI(Z)$  appearing in Chernoff's formula is simply the lowest-order term in an expansion in  $1/N$ :



$$P_N(Z) = \left[ N \frac{d^2 I}{dZ^2}(0)/(2\pi) \right]^{1/2} \times e^{-N(I(Z)+N^{-1}J(Z)+N^{-2}K(Z)+\dots)}. \quad (\text{A17})$$

Now expand each of  $I(Z)$ ,  $J(Z)$ , and  $K(Z)$  as a power series in  $Z$ :

$$I(Z) = (i_2 Z^2 + i_4 Z^4 + \dots), \quad (\text{A18a})$$

$$J(Z) = (j_0 + j_2 Z^2 + j_4 Z^4 + \dots), \quad (\text{A18b})$$

$$K(Z) = (k_0 + k_2 Z^2 + k_4 Z^4 + \dots), \quad (\text{A18c})$$

where the constants  $j_0, k_0, \dots$  are required for properly normalizing  $P_N(Z)$  at each order in  $N$ . [There is no term  $i_0$  in the expansion of  $I(Z)$  because the prefactor  $[N \frac{d^2 I}{dZ^2} \times (0)/(2\pi)]^{1/2}$  in Eq. (A17) ensures that  $P_N(Z)$  is already normalized at lowest order. Of course,  $i_2 = \frac{1}{2} \frac{d^2 I}{dZ^2}(0)$ .] Using  $Y^2 \equiv NZ^2$ , we can then rewrite Eq. (A17) as

$$\begin{aligned} P_N(Y) &= e^{-j_0} e^{-i_2 Y^2} \exp \left[ -\frac{1}{N} (k_0 + j_2 Y^2 + i_4 Y^4) + O(N^{-2}) \right] \\ &= e^{-j_0} e^{-i_2 Y^2} \left[ 1 - \frac{1}{N} (k_0 + j_2 Y^2 + i_4 Y^4) + O(N^{-2}) \right]. \end{aligned} \quad (\text{A19})$$

In this form the correspondence with the Edgeworth expansion becomes clear. The normalization of  $P_N(Z)$  is fixed by  $j_0$  and  $k_0$  to that order, and the  $i_2$  term represents the central limit theorem result, with  $i_2 = (2\sigma_x^2)^{-1}$ , and the  $1/N$  terms are the leading-order corrections predicted by the Edgeworth series. For each term in the Edgeworth expansion, there is a piece that dominates at large  $Y$ . Of

course, this is the term that contains the highest power of  $Y$ , e.g., the term  $i_4 Y^4$  in Eq. (A19). Large-deviations theory can be thought of as a clever way of summing up all these terms to determine the dominant large- $Y$  behavior.

Let us now show how this connection works for the binomial distribution. In that case, the first two cumulants are easily shown to be

$$\kappa_2 = 1, \quad \kappa_4 = -2. \quad (\text{A20})$$

Thus its Edgeworth expansion is given by [cf. Eq. (2.18)]

$$P_N(Y) = \frac{1}{\sqrt{2\pi}} e^{-Y^2/2} \left[ 1 - \frac{1}{12N} (Y^4 - 6Y^2 + 3) \right]. \quad (\text{A21})$$

If one expands the  $N$ -dependent exponential in (A12b) to leading order, one obtains precisely (A21). Now from (A11), we may identify the functions  $I(Z)$ ,  $J(Z)$  and  $K(Z)$  appearing in (A17) as

$$I(Z) = \frac{1}{2} (1+Z) \log(1+Z) + \frac{1}{2} (1-Z) \log(1-Z), \quad (\text{A22a})$$

$$J(Z) = \frac{1}{2} \log(1+Z) + \frac{1}{2} \log(1-Z), \quad (\text{A22b})$$

$$K(Z) = -\frac{1}{12} \left( 1 - \frac{4}{1-Z^2} \right). \quad (\text{A22c})$$

By expanding each of these functions around  $Z = 0$ , we obtain  $i_2 = 1/2$ ,  $i_4 = 1/12$ ,  $j_0 = 0$ ,  $j_2 = -1/2$ , and  $k_0 = 1/4$ . Substituting these values into (A19), we recover precisely Eq. (A21).

- 
- [1] L. Barack and C. Cutler, Phys. Rev. D **70**, 122002 (2004).
  - [2] J. R. Gair, L. Barack, T. Creighton, C. Cutler, S. L. Larson, E. S. Phinney, and M. Vallisneri, Classical Quantum Gravity **21**, S1595 (2004).
  - [3] W. Feller, *An Introduction to Probability Theory and Its Applications* (John Wiley & Sons, New York, 1970), 2nd ed., Vol. II.
  - [4] I. S. Shiganov, J. Sov. Math., **35**, 2545 (1986).
  - [5] V. Petrov, *Sums of Independent Random Variables* (Springer-Verlag, New York, 1975).
  - [6] This argument is taken from M. Bazant, "Asymptotics Outside the Central Region" MIT OpenCourseWare,

2005, <http://ocw.mit.edu/OcwWeb/index.htm>

- [7] D. W. Stroock, *An Introduction to the Theory of Large Deviations* (Springer-Verlag, New York, 1984).
- [8] J. Lewis and R. Russell, "An Introduction to Large Deviations for Teletraffic Engineers," 1996, <http://www.stp.dias.ie/APG/apg.html>
- [9] D. Marković, Phys. Rev. D **48**, 4738 (1993).
- [10] C. Cutler and J. Harms, Phys. Rev. D **73**, 042001 (2006).
- [11] J. A. Edlund, M. Tinto, A. Krolak, and G. Nelemans, Phys. Rev. D **71**, 122003 (2005).
- [12] T. S. Nanjundiah, Am. Math. Mon. **66**, 701 (1959).